ERIC Document Resume	See Instruction Sheet	DATE OF RESUME	DATE RECEIVED

A. PUBLICATION DATA

1. ERIC NUMBER	2. ORIGINATOR CONTROL NO(S).	3. IDENTIFICATION NO.
ED001159	NA	OE-25003

DO NOT WRITE IN SPACE BELOW

4. AUTHOR(S): Guidance, Counseling, & Testing Section, Division of State & Local School Systems

5. TITLE OF DOCUMENT ABSTRACTED:
Understanding Testing: Purposes & Interpretations for Pupil Development

6. PUBLICATION TITLE:
NA

MICROFILM COPY AVAILABLE
YES ☒ NO ☐

COPYRIGHT GRANTED
YES ☐ NO ☒

COPYRIGHT RELEASE FILED
YES ☐ NO ☐

TYPE OF REPORT
PROPOSAL ☐
IN-PROGRESS ☐
FINAL ☒
FOLLOW-UP ☐

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

7. EDITOR(S):
McLaughlin, Kenneth F.

8. PLACE OF ORIGIN
ORG. U.S. Dept. of HEW, Office of Education,
ST. Wash., D.C.
CITY, STATE

9. PUBLISHER OR SOURCE
U.S. Govt. Printing Office
Washington, D.C. 20402

TELE. EXT.
INST/LOC CODE

10. VOL.	11. PAGES	12. MO./YR.	13. TOTAL PP.
NA	NA	63	24

B. INDEXING DATA

14. SUBJECT AREA

15. TYPE OF ENDEAVOR
Report

16. INDEXING TERMS (6-12)

guidance	use of tests		
cumulative record			
test construction	test classification	test administration	
test programs	counseling services	test interpretation	

17. ABSTRACT (MAX. 250 WORDS)

These articles are written to give essential facts on testing to interested citizens. The importance of testing lies not in being an end, but a valuable tool in guidance. Tests represent only a sampling of a person's performance from which general characteristics are inferred. There are numerous types of tests designed for specific purposes: measurement in some is indirect. Conditions under which tests are taken influence results, and accuracy of interpretation determines their usefulness. Test construction demands using advice of specialists and developing administrative procedures, and norms for interpretation.

Classification of tests is given on the basis of testing procedures and purposes. Large scale testing programs are described. Tests are used for identification of ability level and for selection for colleges and jobs. Group classification within schools and evaluation of teaching are aided by tests. They supply data for educational and vocational planning and assist counseling. Caution is needed in interpreting and reporting of test results. The cumulative record serves as the foundation of the guidance program. It is a depository of objective information. Its standards and uses are evaluated.

Programs of testing are important to guidance programs; current programs are described but each school's program should be tailored to meet particular needs. The school staff in its administration of the testing program determines its usefulness. Techniques for selection of tests and utilization of results are suggested.

Definitions of common measurement terms follow the articles.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION	20. DIVISION OR DEPT.	21. CONTRACT NO. / PROJECT NO.
	22. TOTAL COST OF PROJECT	
	23. DURATION OF PROJECT	U.S. Department of Health, Education, and Welfare Office of Education

371.26

Lab

ED001159

OE-25003

UNDERSTANDING

TESTING

PURPOSES AND INTERPRETATIONS FOR PUPIL DEVELOPMENT ★ ★ ★ ★ ★ ★ ★ ★

REPRINTED BY PERMISSION
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN PRINTED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE • Office of Education

371.26

ED00 1159

OE-25003

UNDERSTANDING

TESTING

PURPOSES AND INTERPRETATIONS FOR PUPIL DEVELOPMENT ★ ★ ★ ★ ★ ★ ★ ★

by the Guidance, Counseling, and Testing Section, Division of State and Local School Systems

edited by

KENNETH F. MCLAUGHLIN, *Specialist, Appraisal of the Individual*

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

ABRAHAM RIBICOFF, *Secretary*

Office of Education, STERLING M. MCMURRIN, *Commissioner*

Educational Materials Laboratory
Office of Education

NOV 1964
DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Printed 1960

Reprinted 1962

Reprinted 1963

U.S. GOVERNMENT PRINTING OFFICE

WASHINGTON : 1963

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 - Price 25 cents

CONTENTS

	Page
For Better Understanding, by Frank L. Sievers	1
A Few Basic Facts About Tests, by Paul MacMinn	2
How Is a Test Built?, by Kenneth F. McLaughlin	4
Types of Tests Available, by Robert C. Hall	8
Uses of Tests, by Dolph Camp	12
The Cumulative Record, by Royce E. Brewster	14
Guidance and Programs of Testing, by Carroll H. Miller	16
Administration of the Testing Program, by Frank E. Wellman . .	19
Definitions of Commonly Used Measurement Terms	23

By FRANK L. SIEVERS, *Chief, Guidance, Counseling, and Testing Section*

FOR BETTER UNDERSTANDING

A testing program can be successful with complete and open understanding

THROUGH the following discussion of testing and its place in the schools of the country, the Office of Education hopes to contribute to public understanding of the testing phase of guidance. These articles are not written for test experts, or for other persons familiar with testing practices. Rather, the purpose is to present essential facts on testing for parents and other interested citizens. Teachers and administrators, busy with large enrollments and countless details, may find in the articles information on some current practices in testing.

Of course, it is not possible to present a complete picture of testing in a series of brief articles. Much less is it possible to offer an adequate description of the total program of guidance services. Many phases must be omitted completely, or at best suggested by a hasty sentence or paragraph. Among the phases of guidance omitted are techniques other than testing for studying abilities, interests, adjustments, and goals of the individual; the counseling process; the use of educational and occupational information; and the relation of guidance to other personnel services.

With the scope thus narrowed, the discussion is focused on the purposes of testing in the schools, the science of test construction, the nature of commonly used tests, some of the uses made of test results, the nature and use of cumulative records, and the administration of testing programs. Throughout the discussion there is one recurrent theme: The testing program is not an end in itself, but rather a valuable tool used by school administrators, supervisors, teachers, counselors, and psychologists. Overall responsibility for coordinating the articles was assigned to Kenneth F. McLaughlin, specialist in appraisal of the individual, of the Guidance, Counseling, and Testing Section.

The term "guidance testing" applies not only to selecting and administering tests but also to interpreting and using test results in helping the pupil to realize his potential. Tests provide guidance workers valuable information which, along with other essential facts, is useful in helping pupils select school subjects, achieve satisfying adjustments which will encourage them to stay in school, and make more satisfactory career choices. Tests are particularly valuable in providing objective estimates of a pupil's abilities and achievement when he transfers to another school, continues his education beyond high school, or enters the labor market.

Recently there has been much public discussion of the use of tests as a part of the college admission procedure. Citizens are well aware of the pressure of increasing student populations on our colleges and universities, but pupils and parents of pupils preparing to enter college have an immediate and personal interest in any means colleges use in selecting pupils. Such interest is natural and proper, but it is important that the value of testing as a part of guidance not be obscured by an overwhelming interest in some special application, as, for example, in the college admission process. The broad view, both expressed and implicit in the discussion, should contribute toward the maintenance of a balanced outlook on the relation of testing to the total guidance effort.

The Office of Education is charged with special responsibility for assisting States in providing testing services in secondary schools as authorized under the National Defense Education Act of 1958. Within the Office, responsibility is delegated to the Guidance, Counseling, and Testing Section. The responsibility can be effectively discharged only on the basis of complete and open understanding by all concerned.

GUIDANCE and COUNSELING

The two words are not synonyms. Guidance, the bigger term, includes counseling as one of its 6 elements.

1. Analysis: Helping the student get the facts about himself—from test results, cumulative records, and other means of identifying potentialities and interests.
2. Information: Giving him the facts about his environment—about educational and occupational opportunities and requirements.
3. Orientation: Helping him to get acquainted with the school program and educational and vocational opportunities and requirements.
4. Counseling: Helping him to develop self-understanding and to develop his educational and occupational plans.
5. Placement: Helping him carry out those plans.
6. Followup: Determining how his plans worked out and how effectively the educational program served him.

By PAUL MACMINN, *Specialist in Preparation of Personnel Workers*

A FEW BASIC FACTS ABOUT TESTS

necessary to understanding their uses and value

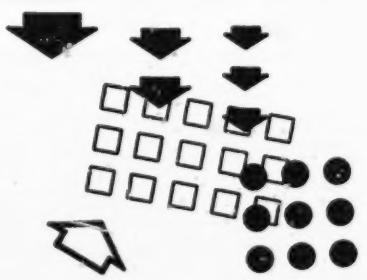
TOO FREQUENTLY the popular concept of the value of tests, especially objective tests, is one of extremes. At one extreme are the people who think that the results of testing are almost infallible measurements of a person's abilities and aptitudes and should be accepted above all other available information. People holding such a view tend to close their eyes to the shortcomings of tests and to overstress their objectivity and usefulness.

At the other extreme are the people who discredit all tests (for example, the ones used to predict vocational potential) as being unreliable and almost useless in understanding and predicting human performance and behavior. In fact, persons holding such a view contend that schools should discontinue using tests because they believe more harm than good comes from their use.

Neither of these extreme points of view is defensible, and each undoubtedly exists because of misinformation, for the facts indicate that the true value of tests lies at some point between these two extremes. Because tests are praised or condemned out of proportion to their true value and because there is much misunderstanding of the school testing program, this article attempts to present some of the facts which should be known about tests.

A Sample of Performance or Behavior

How a person responds to a test is essentially a sample of a person's performance or behavior. This means that tests are designed or constructed to measure representative, or sample, performances or behaviors from which the whole, within reasonable limits, can be inferred, not to measure all possible performances or types of behavior, whether general or specific. For example, a construction engineer planning to erect a tall building wants to know something about the underlying structure of the ground on which the building is to stand. Instead of digging down to the desired depth over the entire foundation area, he obtains sample cores of the earth down to bedrock with a boring device.



With a sufficient number of such samples he is able to get a reasonably accurate estimate of the total underground structure. A similar type of process is followed in sampling with tests. If a teacher wishes to know the extent of a pupil's vocabulary the most accurate way for her to determine it would be to ask him to define all the words in the English language. Such a method would obviously not be feasible, so, like the construction engineer, she would use the sampling process, selecting words of known difficulty from lists of the commonly used words at each age and grade level.

Since tests are merely samples of performance or behavior, great care must be exercised in choosing items for them. There must be a sufficient number of items and they must be representative of, or in some way known to be related to, the performance and/or behavior to be measured. For example, a teacher would learn little about a 7th grade pupil's knowledge of arithmetic by giving him a test consisting of only 5 examples in addition. Even if she increased the number of examples to 500 but confined them to addition, her knowledge of the pupil's arithmetical ability would remain severely limited. To understand what the pupil knows about arithmetic, she would have to test him on each of the fundamental operations and with a variety of problems. His ability to handle the sample items would then reflect his ability to handle most, if not all, of the arithmetical processes such as addition, subtraction, multiplication, or division.

The more complex the human performance or behavior, the more difficult it is to measure. This may seem like a completely obvious observation, but now and then our doubts are raised when we hear people carelessly quoting a child's IQ score with the same degree of confidence as they would use in reporting his weight.

Because of degrees of complexity, certain types of performance or behavior lend themselves to a direct form of measurement, while others must be measured indirectly. For example, the complex function of aca-

demic aptitude (commonly referred to as intelligence) is made up of a variety of elements. Some of these elements such as the extent of a pupil's vocabulary or arithmetical knowledge lend themselves to a more direct form of measurement. Others such as the ability to perceive relationships or to manipulate objects mentally are more complex and must be tapped through an approach which is rather indirect.

We use numerous types of common indirect measurements in our everyday living. For example, we do not measure heat and cold directly but through the expansion or contraction of a column of mercury in a thermometer. In other words, we are not able to observe temperature directly but we estimate it on the basis of its observable effect on the measuring instrument. So it is, in a sense, in some of the items used in a test for the measurement of a complex function such as academic aptitude.

Representative of the Whole

A pupil taking an academic aptitude test is requested to perform certain tasks which have been shown to have some relationship to success in school work. Suppose, for example, that we review some of the tasks which a pupil must perform to be successful in school. We find, among other things, that he needs to read with understanding the kinds of materials found in school books. This implies both an adequate vocabulary and the ability to deal with sentence structure. Again, we find that he needs some knowledge of arithmetic, which involves the basic understanding and skill to apply numbers in problem situations. We can readily understand the reason for measuring these types of abilities.

But, using an example of an indirect approach, an academic aptitude test for younger pupils might also contain, as one of the items, a sketch of a chicken, a pigeon, a cat, and a robin. In the directions the pupil would be asked to pick out the one sketch which was different from the others. On the surface this may seem a rather silly exercise which is totally unrelated to the question of how well the pupil might be expected to perform school work. Without some type of evidence to the contrary, it could well be. If, however, through an experimental process, test constructors had found that the ability to perform the task calls upon abilities which in some way are similar to abilities which make for success in school work, then its inclusion in the test would be justifiable.

To Be Used Wisely

If tests are to contribute to the school program, they must be used wisely and with caution. Persons responsible for a testing program should bear in mind:

- Tests in and of themselves are far from being perfect measuring instruments.
- The attitude and motivation of the pupil taking a test and the conditions under which he takes it influence the results.
- The accuracy with which the results are interpreted determine to a great extent the usefulness of tests.

It should also be realized that there are many types of tests and that each one has been constructed to measure in a specific area for a specific purpose. For example, a personality test (or scale, as it is frequently called) designed for the measurement of certain personality characteristics should not be treated as a test which could predict a pupil's success in college. It is known, of course, that personality characteristics have some bearing on a pupil's chances for success, but if we are trying to assess his mental ability to perform school work, we should use some type of academic aptitude test.

Because there are numerous types of tests designed for specific purposes, tests must be carefully selected for a particular use. One test cannot be taken as a measurement of all of a person's aptitudes, traits, or abilities but must be accepted simply as a measure in the specific area for which the test was constructed.

In summation, we could reply to some of the critics of tests and to the overzealous defenders of tests by paraphrasing Lincoln's sage remark: It's not so much the use of a bad thing but the abuse of a good one. Most tests are well designed and constructed and when used wisely for the purposes intended are instruments which can help us reduce substantially the guesswork in understanding and predicting human behavior and performance.

The importance and usefulness of tests in evaluation and teaching cannot be overestimated. There is a great need for improved and more adequate instruments for evaluation of many kinds of educational growth. The misuse of tests and the misinterpretation of test data continue to be a glaring danger to good educational programs. Any test instrument should be selected and used in terms of what a particular school had intended to teach. To judge a school solely on the basis of data derived from any battery of examinations is an invalid and dangerous venture. State or national examination results, used without due regard for the educational objectives of the school and the nature of the student body, are likely to be misleading.

Resolution No. 13 adopted by the American Association of School Administrators in convention, February 18, 1959.

620768°—62——2

By KENNETH F. MCLAUGHLIN, *Specialist in Appraisal of the Individual*

HOW IS A TEST BUILT?

In answering this question, Dr. McLaughlin describes the careful step-by-step procedure a teacher or publisher should take to construct a valid and useful test.

EACH of you is familiar with such daily tests as sampling a bowl of soup to see if it suits your taste or sighting down several long boards to determine which one is most suitable for a kitchen shelf. While a few tests in school are similar to the "work-sample" type just described, most of them are of the "paper-and-pencil" variety made by the teacher.

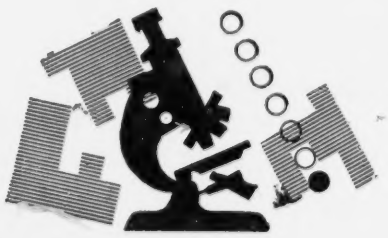
Daily or weekly teacher-made tests serve a definite purpose as indicators of pupil and class progress in specific subject matter. You may remember that many of the subject tests of your school days consisted of three or four essay questions written on the blackboard. As a pupil you spent part of the class period writing essays on the questions—being specific, if you knew the answer, or verbose, if you were not too sure of your subject.

While you may have been unhappy about the amount of writing required by a test, you were never much concerned that after regular school hours your teacher had to read all that you had written to try to evaluate it fairly. Your teacher may not have had a course in tests and measurements during her formal teacher training and consequently may not have known how to construct objective tests which cover more topics in less time and require minutes instead of hours to score.

Most published tests of the objective type are called standardized tests. They are available as single-subject tests or as a battery of several tests covering a number of subject areas. Some objective tests, however, are prepared to be used with certain textbooks and are not standardized.

A standardized test is one which samples the subject matter in a systematic way, is administered in a particular way, is scored objectively, and has meaningful norms for comparing the achievement of pupils.

There are a number of reasons for making standardized tests a part of the evaluation program of a school. First, in a given period of time, a large body of information can be covered. Second, test items or test questions are included which aid the teacher in distinguishing good pupils from poor ones. Third, they make it possible to compare the achievement of each pupil with the achievement of his class or with the pupils of another year, another city, or of the country as a whole.



Steps in Constructing

Described below are some of the procedures a test publisher must take in order to make available tests which will give a reasonably accurate measure of a person's aptitude for learning or the degree of his mastery of a specific subject area at a specified grade level—kindergarten through college. Most of these steps can also be used in constructing a classroom test—although it would not be completely standardized.

To illustrate, suppose a test publisher decides to build a measuring instrument in eighth-grade mathematics. The company's mathematics specialist would procure from many school systems, for grade eight, a number of recognized texts in mathematics and representative courses of study. He would analyze the content and amount of time scheduled for each major topic.

Further, he would seek the assistance of the best available curriculum specialists in mathematics in schools and universities. He might solicit advice of these experts individually or bring them together in a conference to consider the educational objectives to be measured and the topics to be included in the test.

Armed with information from the analysis of current textbooks and the suggestions of the experts, the company specialist would develop a table of specifications outlining the kinds of topics and subtopics to be covered by the projected test. Such a table would correspond to the shopping list a cook makes in planning a dinner or the materials list of a craftsman planning a coffee table.

In using any recipe for a dinner dish or a blueprint for making a table, you must know not only the kinds of materials in it but also the quantity of each. Similarly, to obtain a well-balanced test, the company specialist uses the textbook analyses and the opinions of the subject specialists in determining the relative importance of each topic and the proper proportion of the final test which should be devoted to it.

On some essay tests of your school days, you may recall, the teacher indi-

cated the number of minutes you should spend on each topic or question, depending on its importance. With a standardized test, this time rationing is no longer necessary since the amount of time spent on any one topic is automatically controlled by the number of questions chosen for it.

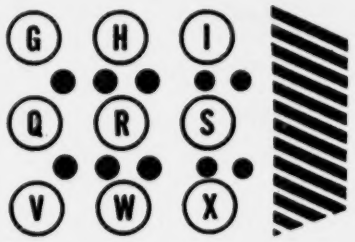
Most tests are designed either to cover one complete class period of 40–50 minutes or to have one or more of the subsections completed within that time limit. Psychological tests, reading tests, and the like are generally long enough to keep all pupils busy during the complete testing time. In most subject-matter tests the time allotment is usually generous enough to permit most of the pupils to try all of the items.

Among the kinds of test items used in an objective test are true-false, matching, or multiple-choice. The type used most frequently is the four-choice or five-choice multiple choice item, although in certain subject areas other types may be more appropriate.

After the proportion of questions for each topic has been determined, a number of qualified mathematics teachers may be requested to write four or five multiple-choice test items on each specified section of the test. The publisher provides special instructions and sample items as guides. He encourages each item writer to use his ingenuity and skill in assembling questions that require the pupil to think or to apply principles based on the material of the course he has studied rather than merely to recall facts.

Item writing is hard work. It takes time to construct a good clear question with four or five choices. While only one choice may be correct, the other choices must appear plausible or reasonable to the pupil who has only partial information, misinformation, or who does not understand a fundamental principle. No test maker attempts to write complicated or trick questions, although for every test there are nearly always some pupils who insist that such questions are included.

The total number of items requested from item writers is generally four or five times the number actually needed. For example, for a 50-item test, at least 250 questions may be constructed. If there are enough usable items, it may be possible for the publisher to make two comparable forms—that is, two similar tests with questions covering the same subject matter.



We should point out, however, that two comparable forms of a test do not have to have exactly the same types of questions—with just a number or a few words changed. Any test is just a *sample* of the larger body of information being tested. From a group of 10 available items on a certain topic the test maker can choose 2 different items for each form of a test in order to meet the requirements of the table of specifications. After the tests are completed and normed, the two forms will be comparable.

Trial Run Made

After the subject-matter specialists have submitted the test items to the company's professional item writing expert, he examines them for editorial defects and classifies them according to the table of specifications. Next, he selects the number of items needed for the test or for each of several forms of the test and assembles the test for a trial run, called a pretest.

In order to conduct a pretest, the publisher must obtain permission from several representative schools at the proper grade level to administer the test. On the basis of this trial each item is analyzed to determine three things:

- The difficulty (the more pupils who choose the right answer, the easier the question).
- The appropriateness of the choices (if no one selects a particular wrong answer, this choice may be replaced).
- The discrimination or separation of the better pupils from the poorer ones (no test maker wants to include items which more poor pupils than good pupils will mark correctly—which is termed "negative discrimination").

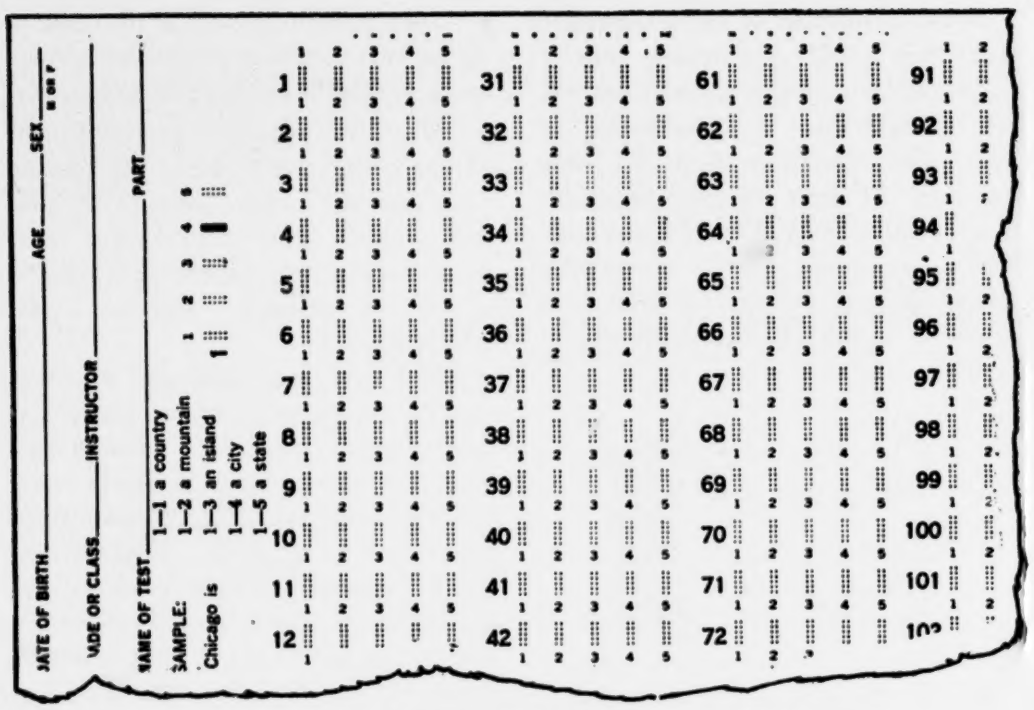
After this item analysis, an item may be rewritten or a new one added to replace a poor one. Then a second trial of the test is made with different schools and the item analysis procedure repeated. Experienced test makers seldom have to carry out the process a third time. However, a test maker must have at least one trial run before attempting the final assembly of a test, for no matter how much experience he has brought to bear, "bugs" may creep in.

Instructions on Administering

While the editing, the proofing, and the final printing of the test are being done, the standardized administrative procedures are developed. These instructions include the exact words the teacher will read to the pupil in conducting the test, the timing of the whole test, and the instructions for scoring.

Below grade four, pupils mark their answers in test booklets which therefore can be used only once. In grade four and above, however, it is possible for the school to purchase, for many kinds of tests, a reusable test booklet and an expendable separate answer sheet—which diminishes the overall cost of a testing program. If a new test is printed in a reusable booklet form, a separate answer sheet must be designed that can be scored quickly by hand or machine. A scoring key must also be planned, prepared, and punched to fit the answer sheet.

For schools in which the teachers must score the tests, several types of self-scoring answer sheets have been developed. One type consists of tamper-proof double sheets with a carbon sealed between. When the sheets are separated, it is easy to count the cor-



A sample 5-choice multiple choice answer sheet

rect answers. A similar type is prepared with chemically treated paper which will clearly record the correct answer or indicate the incorrect answer which was chosen. In another type of test pupils use a pin to punch holes through several thicknesses of paper.

Some large schools or county systems rent test-scoring machines which will score properly marked answer sheets at the rate of over 200 an hour—and thus release teacher time for more profitable tasks than counting dots or marks. These machines can also indicate how many pupils in a class choose each answer for each question. Such information aids the teacher in locating class deficiencies.

One of the most recent developments in the test-scoring field is an answer card which can be scored and punched with a new machine and then sorted on standard machines. Some large-scale programs now use machines which simultaneously "read" the name of the pupil and score all the tests of a battery—at the rate of 5,000 papers an hour.

Norms Developed

The next step for the test publisher is the development of national or regional norms. By this we mean that the publisher must reach an agreement with a sample of representative schools to administer the test to all of the pupils of the appropriate subject group. The schools are chosen from many cities, towns, and rural areas; white collar and industrial areas; upper level, middle level, and lower level income groups. (Statisticians call such a sample a "stratified" sample.)

When scoring is completed, tables of numbers, called "norms," are constructed which make the raw score meaningful and give some indication to the pupil, the teacher, the school administrator, and the parent of the relative ability or relative standing of each child.

Publishers may convert raw scores to an appropriate scale and then construct one or more of several kinds of norms, such as age equivalent, grade equivalent, or percentile. Each system offers a different way of giving meaning to the raw score.

Age-equivalent and grade-equivalent norms give the average scores of pupils of a particular age or grade and are most appropriate for elementary-school years. If a pupil in the fifth grade makes a grade-equivalent score for the sixth grade, it means that he made a score equivalent to that of the average sixth-grader in that subject. It does not mean that the child should be in the sixth grade.

One of the most understandable types of norms is the percentile. This measure indicates how a pupil ranks with other pupils of his grade or age group. For example, a pupil whose score is at the 75th percentile is equal to or exceeds the scores of 75 percent of his classmates; 25 percent of the group did better than he did. A pupil whose score is at the 40th percentile is equal to or exceeds 40 percent of his classmates and is just a little below the average, or the 50th percentile.

It should be pointed out that a standardized test for a specific grade is designed so that the average raw score, or number of correct answers, will be about half of the total number of questions. Thus, the old idea of "70 percent is passing" no longer holds with these standardized tests.

Many local norms—they can be easily constructed for a school, city, county, or State—give additional information and are more meaningful than national norms. Such local norms make it possible to compare a pupil with others who live in a known community and have a known background.

A Fair Question

Parents, counselors, and teachers often ask: "How dependable are these tests?" It is a fair question. Dependability might well be discussed in the terms which most manuals now use: "reliability" and "validity." By *reliability* we mean *consistency*. For example, if a pupil scores at the 30th percentile on the test today, may we be assured that he will score at a similar percentile on the same test or another form of the test tomorrow? If the test is reliable, the same pupil would not score at the 75th percentile if he took it again soon without further study, nor even at the 50th percentile; he would score near the 30th percentile.

Because there is likely to be a small shift of percentile points, the current trend is to interpret test scores as a *band* of several percentiles. In other

words, a score at the 30th percentile may mean that the pupil's "true" score may lie somewhere in the band between the 27th to 33d percentile. The width of this band will vary with different tests and will often vary at different percentile points on the same test.

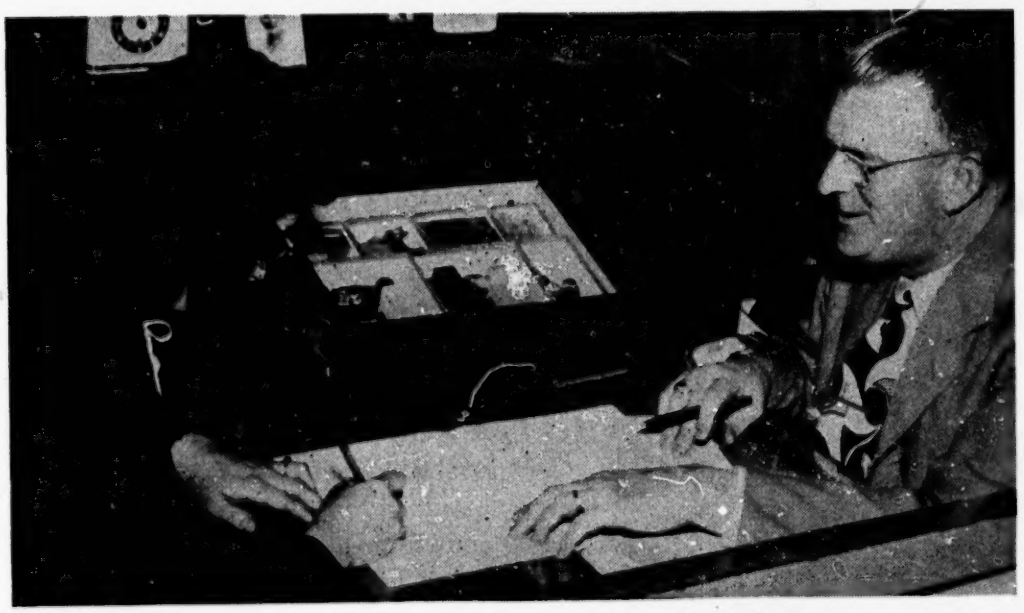
By "validity" we mean "How well does the test measure what it is supposed to measure?" A test maker checks validity by comparing the new test with other well-accepted tests if any exist, or by obtaining the consensus of experts in the subject. It is the responsibility of the test maker to conduct research to learn whether a test is valid for the purposes claimed and whether the scores of the test are reasonably stable or reliable.

Suggestions for the interpretation of the test results are sometimes included with the printed instructions for the test administrator. At other times, a separate manual of interpretation is provided. Such suggested interpretations are one of the most important contributions of the test author, or test publisher, to the proper educational use of the results. They help to make the test scores meaningful to the counselor, the teacher, the pupil, and the parent. Test manuals often include not only an explanation of the norms but also suggested remedial measures for pupils with low scores.

Schools or teachers selecting tests should not rely on the test title alone, but should always examine a copy before placing an order. For example, in achievement testing, it has been found helpful to have a test committee of a school compare several tests in a subject area in order to determine the one test which best suits the current curriculum and which provides the most helpful information.

From the procedures described here, you will see that a standardized test is the result of many weeks of work by many specialists. If the instructions for the test are carefully followed, the results will be meaningful to all.

For a school to obtain testing instruments constructed with care requires only a small amount of the educational per-pupil cost per year. The national average total expenditure per-pupil for education in 1958–59 was $456. A school that has one standardized test battery per pupil per year, administered by the teachers and centrally scored, would probably pay less than one-fourth of 1 percent of the per-pupil cost for its program—approximately $1. Such a percentage of an operating budget does not seem to make cost a major consideration as a school evaluates its needs and programs.



Administration of the Binet Age Scale in an individual testing program, Edward Little High School, Auburn, Maine.

Getting Ready for School

NOT LONG AGO a mother telephoned a school and requested a copy of a specific reading-readiness test. The school did not have a copy of the test available for parents. In an effort to be helpful, the counselor at the school questioned the mother concerning her reason for needing it.

The woman said she understood that local regulations required her 6-year-old child to take a placement test before he entered the first grade. To prevent his being placed in a readiness class for a year, which was the custom in the school system in her town, she planned to coach him.

This made the counselor feel he should explain the reasons for giving the test. A child is tested before he enters school to discover his general ability level, to try to determine whether or not he is mature enough to begin first-grade work, and to decide whether or not he is able to acquire successfully the necessary skills for a beginning reader.

The counselor also pointed out that the test administrator would be skilled in putting a child at ease in strange surroundings, that every effort would be made to obtain a measure of the ability of the prospective pupil, and that, in fact, the standards for these tests had been determined by giving the tests under similar conditions to many children of the same age.

Coaching with the test would be like having a person memorize an eye chart before having his eyes examined. With preknowledge of the questions, a child might well be placed in a class above his maturity level, which would be a disservice to the child. Further, there was absolutely no stigma connected with his being placed in a readiness class, and the child himself would be happier and more apt to succeed in the following grades if he were placed with a group at the same stage of development.

The mother responded that for the first time she understood the usefulness of these tests. With such testing, perhaps her older children could have avoided certain school problems. While she wished to use any legitimate means to help her youngest child do well on his first test, she was willing to let him meet the test situation uncoached.

By ROBERT C. HALL, *Specialist in Program Organization*

TYPES OF TESTS AVAILABLE

Schools are using many types of tests for a variety of purposes. Dr. Hall points out distinguishing characteristics of a number of them.

THERE are many indications that more and more emphases are being placed on tests. Among them are the provisions for guidance, counseling, and testing in the National Defense Education Act of 1958 and the increased use and influence of College Entrance Examination Board tests and the National Merit Scholarship Qualifying Test.

Because of the magnitude of recent testing programs, it is easy for us to forget that tests have been a part of human culture for centuries and to regard them as new and revolutionary. As far back as the first century B.C., China was using a well-developed set of civil-service examinations. In the early 19th century students in Massachusetts faced the demands of oral school examinations. By the middle of the century the enrollment in Boston schools had become so great that there was no time for individual oral examinations and uniform written examinations were adopted for the first time.

Binet in 1904 introduced the first major change in testing by designing an examination which could be used in assessing a child's ability to do school work. Large-scale testing of groups became widespread for the first time during World War I as a method of classifying military personnel. Differential tests of abilities and aptitudes were developed early in World War II. Since that time there has been much research to discover the usefulness of differential tests in the guidance of boys and girls.

Because many kinds of tests have been developed for a variety of purposes, specialists in measurement and evaluation have evolved a number of useful classifications of tests. For the reader unfamiliar with the kinds of tests available, we briefly classify them in two ways: according to the testing procedure and according to measurement objectives.

Classification Based on Testing Procedures

Some of the types of tests classified on the basis of the testing procedure are described below.

Performance vs. Paper-Pencil Tests

Performance tests require the examinee to manipulate objects, for example, to assemble mechanical parts, place pegs in holes, or perform on a typewriter. *Paper-pencil* tests require the examinee to write paragraphs or sentences or mark an answer sheet in accordance with instructions.

Objective vs. Subjective Tests

Objective tests require the person scoring the test to exercise little, if any, judgment. An answer key is prepared when the test is constructed so that no matter how many persons grade the test, the pupil will receive the same numerical score from each. Objective tests usually require the examinee to write a word or check a correct answer. The types of test items employed usually include multiple-choice, matching, true-false, and completion. By using an objective test a large amount of subject matter can be sampled in a class period.

Subjective tests require the scorer to use considerable judgment in evaluating the examinee's responses. Common examples of subjective tests are the oral and essay examinations. Generally, in an essay test the number of questions which can be asked in a class period is limited, since the pupil is generally required to "discuss" each question at length.

Standardized vs. Teacher-Made Tests

A *standardized* test is administered and scored according to specific instructions. Norms are available which make it possible to compare an examinee's performance with that of others in a similar population group. *Teacher-made* tests are those constructed by a teacher, usually for her own use for an informal evaluation of classroom learning. They are made for a particular class which has been taught in a certain manner.

Individual vs. Group Tests

Individual tests are administered by a qualified examiner to one examinee at a time. The examiner records the answers and observes the examinee while he is responding to oral questions or performing the tasks assigned. Among the best known individual tests are the Stanford-Binet Scale, Rorschach Ink Blot Test, Thematic Apperception Test, and the Wechsler Intelligence Scales. *Group*

tests, as the name implies, are those that can be administered to more than one person at the same time, for example, to a whole class. A group test saves time in administration of the test.

Speed vs. Power Tests

Speed tests are those which require the examinee to complete as many test items as possible in a specified time. If time were not the controlling factor, an examinee could complete all the test items. The main purpose is to determine how rapidly the responses can be made.

In a *power* test the examinee demonstrates the extent of his knowledge or depth of his understanding with the time factor eliminated as much as possible. There is, of course, a practical time limit which is determined, in the case of a standardized test, by the amount of time required by most of the examinees in the standardization population to complete the test. Questions are presented in the order of difficulty, from easy to hard.

Classification Based on Purposes of Measurement

Tests are classified in terms of the purpose for which they are designed, for example: mental ability, intelligence, or scholastic aptitude tests; aptitude tests; achievement tests; interest inventories; and personality inventories.

Mental Ability, Intelligence, or Scholastic Aptitude Tests

Mental ability, intelligence, or *scholastic aptitude* tests provide an estimate of an individual's ability to learn by requiring him to perform mental activities that are regarded as evidences of intelligence. Such tests emphasize his knowledge of vocabulary and his ability to reason.

Some tests in this category yield a single total score and others yield separate scores based on both verbal and nonverbal questions. If all that is desired is an overall measure that will aid in the rough grouping of pupils or in estimating success in school, a single test score on a mental ability test will suffice. In most school situations, however, a test which provides both verbal and nonverbal (for example, numerical) scores is valuable as the results may be used in helping pupils make vocational and educational plans.

Aptitude Tests

An *aptitude* test is used in estimating future success of a person in various occupations or further education. In general, such tests attempt to measure certain stable, unitary, and relatively independent psychological characteristics of the person which are believed to be requisite to success on a job or in further education. Examples of these psychological characteristics are memory, spatial perception, word fluency, rhythm, and verbal reasoning. (Scholastic aptitude tests are usually classified separately and are discussed in the preceding section.)

Some aptitude tests, such as those for mechanical, clerical, musical, or stenographic aptitude, attempt to identify and measure the combinations of traits or specific aptitudes believed to be associated with them. For example, mechanical aptitude is believed to include, among others, spatial perception, and mechanical reasoning ability.

The general term "multifactor tests" is used to describe an aptitude test battery or group of tests designed to measure several relatively independent abilities, such as intelligence, memory, verbal comprehension, muscular coordination, and creativity. The tests include a series of verbal and nonverbal sections which measure psychological traits believed to be relatively independent and of varying degrees of importance in estimating job and educational possibilities.

Research has provided considerable evidence that multifactor tests, which have come into use within the last 10 years, are useful in many school situations, for example, in predicting future performance on College Board Examinations and achievement in school subjects.

Achievement Tests

Achievement tests are designed to determine the amount of progress made toward the achievement objectives of the school's instructional program. Some of the objectives of the school are to help the child acquire skill in reading and computation, to learn to think critically, and to help him achieve to his fullest extent in subject areas such as social studies, science, and mathematics.

A special kind of achievement test battery is called a *survey* test. It is used in assessing a pupil's growth in broad subject-matter areas.

Some achievement tests, called *diagnostic* tests, are designed to enable a teacher, counselor, or administrator to determine the pupil's specific strength or weaknesses in such subjects as reading, arithmetic, and language. Test items are grouped according to the academic skill they purport to measure. Frequency of errors in groups of test items identifies the learning needs of the pupil.

Some confusion exists concerning the difference between aptitude and achievement tests. Two kinds of distinctions should be made: The first between the basic nature of aptitude and achievement; the other, between the testing of aptitudes and the testing of achievement.

Perhaps a reasonable distinction between aptitude and achievement might be made in this way: Aptitude refers to the innate ability to achieve; achievement refers to the knowledge and skills that have been acquired. For example, finger dexterity, speed of perception, and spatial visualization are not taught but may be important in some subjects or jobs; skills in arithmetic or reading are

developed to different levels in the classroom.

The distinction between the testing of aptitudes and the testing of achievement is that the former attempts to predict capacity and future performance, while the latter measures the acquired information and skills. It is possible, however, for the test content of an aptitude and an achievement test to be quite similar. For example, a sixth-grade arithmetic test might be used as an aptitude test to predict the achievement in arithmetic or science in the following grades or as an achievement test for sixth-grade arithmetic. In this case there is only a difference in purpose.

Self-Report Inventories

While there are numerous ways of evaluating the personal-social adjustment of the pupil, such as anecdotal records, rating scales, attitude scales, social distance scales, and self-report inventories, here we will discuss only two of the self-report inventories. Although they are not tests in the sense of measuring subject matter which has been learned in the classroom, interest inventories and personality inventories are often called tests. Responses to situations are recorded in a booklet or on a separate sheet.

Interest Inventories

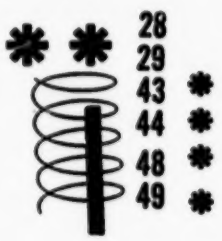
An *interest inventory* is designed to permit a person to record in an organized manner his likes and dislikes, or preferences, in a number of different situations, such as a liking for occupations, choice between activities, or reactions to peculiarities of people. A person's choices are then compared with the choices previously made by successful people in the various occupations, such as physician, production manager, office worker, and sales manager. Or, an inventory may be scored to indicate a larger interest area, such as computational, persuasive, musical, or social service.

It should be pointed out that an interest score is no indicator of the ability of the person, although a person may have a greater chance for success in work which he likes to do. Research conducted thus far has shown that there is considerable stability of interests after grade 10, which suggests a level after which this type of test might prove most useful in helping a pupil to make a vocational choice. When it is used earlier, one should remember that interests may be subject to rapid change.

Personality Inventories

A *personality inventory* is designed to measure personal adjustment, to describe a person in terms of his personality traits, or to discover the nature of his problems. Such tests differ in purpose, types of items, and the traits they purport to measure.

Many educators question whether personality inventories should be considered as a part of the school's general testing program. The decision on the usefulness and appropriateness of this type of inventory must be made by persons in charge of the testing program. Data on the validity and usefulness of personality inventories in school are complex and confusing, but they suggest that such instruments are most helpful when used in conjunction with other data which may be available in the cumulative record. Much care should be exercised in the use and interpretation of results, for much damage can be done when inadequately trained people interpret inventory scores.



Large-Scale Testing Programs

At present, a number of optional testing programs are available to the high school student on a nationwide basis from nongovernmental agencies. Others are being planned. Some of these programs are used in screening high school students for entrance into college, others in determining scholarship awards. The cost of these programs is kept to a minimum. Practically every high school may receive, or can request and receive, printed announcements of them. However, for those persons who do not have access to detailed information, the following descriptions of some of the current programs may be helpful.

National Merit Scholarship Qualifying Test (NMSQT)

The *National Merit Scholarship Qualifying Test* consists of two parts. Second semester juniors and first semester seniors may take the first test sometime in the spring. Announcement of the specific date for the test is posted by high schools early in the school year. Any high school may give the NMSQT by applying for the test. Pupils scoring high enough on the preliminary test must take a second test the following December in order to be eligible for a National Merit Scholarship. Since scholarships are assigned on a proportional basis to each State, it is possible that in some States some of those whose scores are at the 99th percentile may not qualify for the second test. The names of students who do well on the tests are printed and distributed to colleges and universities which may offer scholarships to the runners-up.

Scholastic Aptitude Test (SAT) of the CEEB

The *Scholastic Aptitude Test* of the College Entrance Examination Board (CEEB) is provided to determine the qualifications of high school seniors who desire to enter certain institutions of higher education, either publicly or privately controlled. Students who take the SAT designate the colleges to which the test results are to be sent.

The Scholastic Aptitude Test, requiring a half-day, is administered at regular intervals at designated test centers in every State and in many foreign countries. Information on the dates for the test and the location

of test centers is available from high school principals and guidance counselors.

Many colleges and universities participating in the College Entrance Examination Board program require the student to take not only the SAT but also as many as three subject-matter achievement tests for purposes of academic placement. These tests generally are given in the afternoon of the same day as the SAT is administered.

Preliminary Scholastic Aptitude Test (PSAT)

The *Preliminary Scholastic Aptitude Test*, one of the more recent tests available to high school juniors and seniors, replaces the Scholarship Qualifying Test (SQT) which was administered in many high schools in October of 1956, 1957, and 1958. The PSAT serves two functions: (1) It provides an aptitude test for the guidance of college-bound students early in their junior year, and (2) it provides a screening test for certain scholarship sponsors.

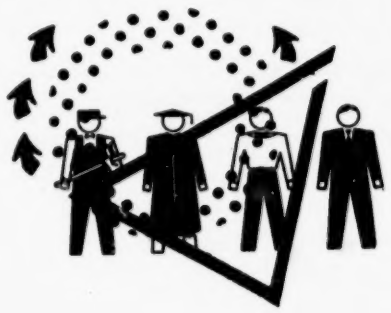
National Secondary School Admission Test (SSAT)

As the name implies, the *National Secondary School Admission Test* is used primarily for screening students seeking admission to grades 7 through 12 in private, independent, and preparatory schools. Schools using this test in selecting applicants would indicate the requirement and availability of the test in their catalogs. The test evaluates general school ability and reading.

Junior Scholastic Aptitude Test (JSAT)

The *Junior Scholastic Aptitude Test* is designed to test the academic fitness of pupils who have finished the elementary grades to continue successfully in the secondary grades. It is used by many schools as a criterion for determining admission or placement of pupils, for general diagnosis of academic strength and limitations, and for the granting of scholarships.

(None of the large-scale testing programs listed are intended to take the place of testing activities existing in public schools. These special programs are limited in purpose in that they are all directed either toward placement in secondary schools or colleges and universities. A school's testing program should continue to evaluate the outcomes of instruction at all grade levels and assist young people in making curriculum and vocational choices.)



Tests of Professional Aptitude

Tests of professional aptitude are designed as an aid in the selection of capable students rather than as a direct aid in the counseling of students. Their main purpose is to serve as a selection instrument for professional schools such as schools of dentistry, medicine, pharmacy, education, engineering, business, law, and nursing. A few tests of this type are available for general use; but most are restricted to member professional schools and official testing centers. States require students who have successfully completed their training in a professional school to pass a licensing examination before permitting them to practice a profession.

Federal Government Tests

The Federal Government does not require or sponsor any kind of a testing program in the schools of the Nation at any grade level. However, the Government does recognize the usefulness of tests and as an employer constructs tests to serve its specific needs. Two such tests are described below.

General Aptitude Test Battery (GATB)

The *General Aptitude Test Battery* was designed for use with adult employment applicants and high school seniors in need of counseling in connection with their registration at the offices of the Federal-State Employment Service. It serves several purposes: To identify aptitudes, to aid in job choice, and to assess vocational strength and limitations. The test may be administered, if needed, to those who register with the Federal-State Employment Service. Administration, interpretation, and scoring of the test battery are handled by the employment service.

Federal Service Entrance Examination (FSEE)

The Federal Service Entrance Examination is handled exclusively by the U.S. Civil Service Commission. It is designed for candidates with potential for becoming top administrators, technicians, and specialists. It is used to fill a variety of Federal Government jobs at several grade levels—most at the starting professional level, GS–5, but some at GS–7 and a very few at GS–9. Under a new plan, the Civil Service Commission will administer these examinations several times during the school year.

To keep the list current, the Commission annually revises its register of persons who successfully complete the examination. The test is open to college graduates and to students who are not below the junior year in college. Announcements of the dates for these examinations are posted in all first and second class post offices and in most college libraries.

AN EFFORT has been made to indicate and classify the large numbers of different kinds of tests which are available—aptitude tests, achievement tests, interest inventories, and personality inventories. Several large-scale testing programs have been mentioned and briefly described. However, no matter how many hours a pupil may spend in taking numerous tests of all types, his time is wasted unless a trained counselor or teacher carefully evaluates the results and uses them along with all other information in the pupil's cumulative record.

By DOLPH CAMP, *Specialist in Testing Programs*

USES OF TESTS

Tests may be used in many ways. Here Dr. Camp discusses six major uses, emphasizing those of value to teachers, administrators, and school guidance counselors.

EACH year millions of standardized and other tests are administered to pupils at all grade levels in the United States and to adults seeking employment or advancement. Unfortunately, in some schools the test results are not being used. Some school officials apparently believe that the test results are "top secret" information since they lock them away in their files. Others apparently consider the results merely as a necessary part of a pupil's official cumulative record—to be entered in his folder in the office, but not always made easily accessible to his homeroom and other teachers. In an increasing number of schools, however, administrators, teachers, and—in schools fortunate enough to have them—trained counselors are using the test results to good advantage.

Tests can be used in many ways. This article discusses briefly only six of the major uses: Identification, classification, selection, evaluation, planning, and adjustment. Even though classification and selection may be considered as subsections of identification and adjustment as a subsection of planning, each of these uses is of sufficient importance to merit independent treatment here.

Identification

Standardized tests may be used, along with other data, to identify pupils with superior talents and those with low ability. For each of these groups, as well as the groups with average abilities, multiple aptitude tests may be used to identify the particular types of abilities each pupil possesses. A pupil with superior talents may be good in English, much better in mathematics, and still better in science, and one with low ability may be weak in history, weaker in English, and still weaker in mathematics. Whatever the pupil's abilities are, a good multifactor or multiple aptitude battery (see p. 9) will depict their pattern.

Classification

Some schools practice homogeneous grouping of pupils; that is, they assign pupils to classes on the basis of mental ability, achievement, test scores, or earlier class grades, because they believe that better instruction is possible if there is a narrow range of abilities among the pupils in a class. Such arrangements prevent any one level of students from setting the pace for the class. If the slow learner sets the pace, the superior student wastes time and loses interest. If the superior student sets the pace, the slow learner becomes discouraged because he finds it impossible for him to keep up. Finally, if the average student sets the pace of instruction, the fast learner is bored and the slow learner discouraged. For the instructional program based on a narrow range of abilities among the members of each class, standardized tests may be used along with other data in classifying students into categories based on different levels of ability.

Selection

Tests may be used along with other data in selecting pupils for a college or university, filling a position in industry, or assigning persons to specialized positions in the Armed Services. Basic to the operation of selection is the fact that there must be more people than there are places to fill. The educational institution, the industry, or the Armed Services set the requirements and select the test in order to determine those who meet, or most nearly meet, the requirements.

The high school or college counselor uses test results in helping a pupil make plans or adjustments. He begins with the pupil's abilities and personality traits and helps select situations in which such abilities and personality traits can best function. For example, if a pupil prefers small classes and likes to know all of his group, perhaps he would be more successful and happier at a small college rather than at a large university.

The personnel director of a steel mill, for example, must find men with certain aptitudes or the skills required for a specific job. In other words, since the requirements of the job are of utmost importance, he looks for persons who possess the qualities it requires.

The problem of fitting employees to positions is different for the Armed Services. Here, the number of persons available is known, the number of positions to be filled is known, and all persons in the Service must have an assignment. From a large group of men and women with varied backgrounds, either volunteers or draftees, the Service must match each person to the job for which he is best suited or to the job where he is most needed at the moment, depending on his particular combination of skills.

With large numbers of men and women to be considered, test results are helpful in matching talents with job needs or as a basis for transfers.

Evaluation

Standardized tests may be used in evaluating instruction. Through the use of test results pupils, classes, and even schools may be compared with fixed standards. Through tests, the level of mastery of essential content by a pupil may be compared with the level which has already been established for his grade or age group. Such a comparison shows whether the pupil is underachieving, achieving beyond his expectancy, or achieving according to his potential.

Through the use of test results a class average may be compared with the averages of preceding classes in order to determine whether changes in methods of instruction are having the desired effect.

Occasionally a principal will find it helpful to compare the average achievement of his school in various subjects with the averages of other schools in his geographical area, whether city, county, or State. However, in the tables of norms, all schools lose their identity so that no school may say that it is better or poorer than the schools of any other particular school system. This is as it should be, since there are too many variables which could account for the differences in average achievement between any two particular schools.

Planning

Perhaps the most important use of tests is that of supplying basic data to pupils, counselors, and parents to aid them in educational and vocational planning. Almost every college or university maintains a profile of abilities a student should have to succeed in its various departments or schools. College officials can compare an applicant's profile with its standard profile to determine in which department or school he is likely to succeed or whether he is likely to succeed in any of them.

A similar use of tests may be made by the industrial counselor who attempts to assist a youth or an adult in obtaining suitable employment, or by a recruiting officer in the Armed Services assisting a new volunteer in choosing a military specialty in line with his abilities and interests and the current needs of the Services.

Adjustment

An important use of tests is in assisting people—children, youth, or adults—in making necessary adjustments. For this purpose test results are used by such people as counselors in educational institutions, personnel managers in industry, and psychologists in clinical counseling. For example, a pupil may be attempting courses of study too difficult for him, a man who is at present a plumber may not be in a position requiring his abilities and interests and needs to change jobs, or a housewife may become frustrated with everyday problems confronting her. In dealing with these people and helping them make adjustments, counselors and clinicians find standardized tests useful.

Cautions in Using Tests

Used properly, standardized tests are valuable tools in educational programs, in industry, in the Armed Services, and in psychological and psychiatric clinics. But, some caution should be observed in their use.

Those administering standardized tests must follow the directions in the test manual exactly. If they do not, the test loses its effectiveness for it will be impossible to make valid comparisons of the test scores obtained with the published standards.

No one should depend on tests alone in making any important career decision. Test results should be used along with such background and educational information as that included in a cumulative record (see p. 15). School marks, data from questionnaires, rating scales, autobiographies, and anectodal records yield valuable information for counseling.

Caution should be observed in the interpretation of test results. It is difficult to set an exact cutting score (that is, a minimum score which must be obtained in order to be considered) as a requirement for entrance into an educational institution, a course of study, or a job. Instead of an IQ or a percentile score, it is usually considered better to use a range of IQ scores or a percentile band (see p. 6–7) in making decisions on requirements.

Caution should be observed in the reporting of test results in schools. Counselors and other test users should have a well-formulated policy on reporting scores. In particular, the policy statement should answer these and other questions: Should parents be given the test scores? Should a pupil be given his own scores? Should all teachers be given all the scores? Although these questions are not discussed here, they should be considered in reporting test results, for all users should exercise maximum caution to see that the persons concerned are helped instead of hurt as a result of testing.

Tests should be safeguarded. The reputable test publisher does not sell his tests to unqualified or unauthorized persons. The use of good tests can be impaired by negligence in the local handling of them. When students or others have studied a particular test in advance, its measuring value has been lessened and the results are often worthless. There are not many closed tests, and tests available on the open market must be safeguarded by persons administering them.

TODAY, schools are overcrowded, there is a shortage of well-trained teachers and counselors, and schools are assuming greater responsibilities for pupil welfare. Many schools are administering tests on a gradewide or schoolwide basis and using the test scores in conjunction with other information for a number of such useful purposes as have been discussed. However, test users must be cautioned that tests are a means to an end and not an end in themselves.

By Royce E. Brewster, *Specialist for Guidance Practices*

THE CUMULATIVE RECORD

The foundation of the guidance program, a depository of objective information

SCHOOL guidance services tend to be effective in proportion to the completeness of the cumulative records that are maintained on each pupil. Effective records show, objectively, the pupil's strong and weak points—the peaks and valleys of his physical, mental, personal, and social capacities.

A Permanent Scoreboard

Although information on a pupil may be gathered by means of many different kinds of questionnaires, written observations, and teacher reports, cumulative record cards or folders permit the consolidation of a large number of test results and personal-social data in an organized form. A pupil's record grows as the pupil grows and provides a permanent scoreboard that should follow him from year to year throughout his school career.

Any discussion of guidance services must consider counseling as the heart of a guidance program. Since each person is unique, his needs and problems should be considered with full recognition of his individuality. However similar a group of individuals may appear, their innate differences vary greatly. The principle of individual differences lies at the very foundation of the guidance program. That is, a teacher or counselor cannot follow a fixed pattern for all students but must tailor the assistance to each.

If a guidance program is to be successful and the results of the testing program are to be used effectively, then the school must look to an adequate record system. The unique value of a cumulative record lies in two facts: The record of any trait of an individual over a period of years is more significant than the record of that trait taken at any one point; estimates of many different traits afford a much more accurate picture of the pupil's educational and vocational possibilities than the estimate of a single trait.

The cumulative record, then, is essentially one of the tools in the individual case study method. It is a depository for extensive objective information about a pupil which may be used by the counselor or teacher as a source of facts, not opinions. Although some of the information may be subjective, the data compiled over a period of years will provide a composite picture of the pupil. The record will have validity because it represents the pooling of information from the school officials having direct contact with the individual pupil. The process will tend to minimize the influence of opinions of a single counselor or teacher in dealing with pupils with specific problems.

A cumulative record is as necessary to all kinds of personnel work—school or industry—as the complete health record is for treatment of a patient by a physician.

Of Facts Needed

Because the cumulative record can help the school facilitate the growth of a pupil in the light of the accepted educational objectives, it should carry the kinds of information teachers and others need to attain this end. The good cumulative record does several things:

- It reveals the rate and trend of development in various intellectual, physical, social, and character traits.
- It is a tool in the program of individual counseling of a pupil and helps all school officials who come in contact with him to work more intelligently with him.
- It is useful as a guide in the revision and development of the instructional program and the curriculum.
- It provides a means for appraising the effectiveness of the school's educational program and the progress of an individual pupil at all levels of school structure and pupil attainment.
- It serves as a formal record of the school's effort to discover and provide for the individual differences of pupils.
- It preserves a cumulative picture of the individual's growth in these traits and makes this picture available to educational staff members as they work with an individual or his parents for any purpose which would help the pupil or assist in promoting his adjustment to a situation or in solving a problem.

The cumulative record should begin with information about the pupil before he enters school and continue until he has been out of high school at least 5 years. Such a long record will show how well his school served him, either in preparing him for college or in preparing him for employment. When kept on the pupil who goes to college, the record of his 5th year out of high school will show his adjustment to college and to his first job. When kept on the pupil who goes into employment on graduation from high school, or as a dropout, information about him for 5 additional years, showing how well he has adjusted to his first and successive

jobs, may be useful to the school in its curriculum planning.

For Pupils, Parents, and School

It is essential that the record follow the pupil in his school progression. The data on the record and the record itself belong to the pupil and to the teacher, counselor, or other school staff member who records the data and who is also responsible for helping the pupil understand and interpret the information recorded.

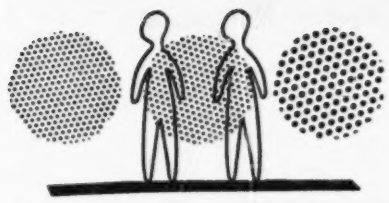
Cumulative records should be available and accessible to all school staff members who have direct contact with the pupil.

The first teacher to whom the pupil reports at a basic station or homeroom should either have a copy of his cumulative record or a summary of the significant facts. The counseling staff and special school staff members such as nurses, psychologists, visiting teachers, and attendance officers should have access to the records. In turn, they may be charged with responsibility, within their functional areas, for supplying such data as may be valuable in dealing with pupil problems.

The record must be factual insofar as possible. Opinions that do not represent the composite subjective opinions of several teachers on various items should be used with caution.

Only necessary information that the school personnel can and will use should be kept. It is imperative for us to remember that a good record system *does not* insure that *wrong* data will not be written down, that information will be kept up to date, and that the record will not be used for unprofessional reasons. Though such misuse is unlikely, teachers, counselors, parents, and others using the records are human beings and subject to errors of judgment.

The data in a cumulative record system should be available only to school staff members working with pupils and their parents on any cogent and immediate problem; ordinarily it should not be available to other persons as it is regarded as a privileged communication. The essential information should be made available in some form to colleges, placement agencies, or employers on request.



Covering Specific Items

Each school system has studied or should study its own needs in order to determine the kinds of information on the pupil which would be useful in helping him. However, a specific group of items should be included in every set of school records. Some of them are given below.

(1) *Autobiography, including family background, home and neighborhood environment.*—The family and cultural background, as shown by the autobiography, has considerable bearing on the pupil's immediate achievement in school and his future education and vocational plans. From such information teachers and counselors may often gain insight of great value in working with the pupil and his parents.

(2) *Health history, including physical and medical examination data.*—Properly evaluated, such material may have a direct bearing on the pupil's school progress and future plans. For the evaluation of such data, the school teaching and guidance staff and the school nurse depend on the school physician or the pupil's physician.

(3) *Standardized test records.*—Many records of test data are incomplete in the information necessary later to identify the test used. Space should be provided for the name of the test, form, grade level of the test, norms used, date administered, and any special conditions that might affect the results. Marked answer sheets or test booklets are sometimes helpful. With complete records, trends in class achievement can be observed and studied. For example, teacher's marks on essay-type tests and class work run consistently higher for girls than marks for boys, but scores on standardized tests usually reverse the direction of superiority—a finding that emphasizes the need for both kinds of data in cumulative record files.

(4) *School marks.*—Class marks and standardized test results are not to be regarded as duplicative. Marks often reflect the presence and the influence of many personality factors of educational significance that impartial and objective standardized tests do not measure.

(5) *Anecdotal items.*—Short descriptions of observed behavior supply evidence of character, personality, attitudes, and interests. No interpretation is given. A number of these on the same student can be helpful to a record user in spotting difficulties.

(6) *Participation in cocurricular activities.*—A pupil's participation in cocurricular activities should be recorded, such as music and dramatic clubs and athletics. Positions of leadership as indicated by offices he has held should also be included.

(7) *Work experience.*—Such work as delivering papers, baby sitting, clerking in a store, or other employment and helping in church bazaars and in scout organizations should be included.

THE CUMULATIVE RECORD properly kept is a source of information. The value of the record largely depends on the adequacy of the information on the individual: His social and economic background, occupational experience, recreational activities, physical condition and health, personality traits, tests, and other significant information. The interpretation of such data is used as a basis for counseling and for predicting future success, inasmuch as educational achievement in the past is perhaps the best guide to the future.

Guidance for an individual pupil should result in his having as good a chance as the school can give him for becoming educationally prepared, occupationally competent, and personally and socially adjusted for happy and successful living.

By Carroll H. Miller, *Specialist in Counselor Preparation*

GUIDANCE *and* PROGRAMS OF TESTING

No one program can be recommended for all schools; each school should tailor its program to meet its particular needs.

TESTING is as old as schools themselves, and tests have a variety of purposes. Here, however, we are concerned with testing programs in relation to guidance services.

In the context of guidance we think of tests as one among a number of tools useful in assessing a pupil's abilities and interests. Now, of course, a tool can be misused, but misuse is scarcely a defect inherent in the tool itself. Probably everyone has at one time or other mashed his fingers with a hammer without suggesting that hammers be removed from tool kits.

Current Programs

Citywide testing programs were developed in some of the larger cities a number of years ago. Then the idea of statewide testing programs developed and spread. By 1955 Segel (Office of Education, Circular No. 447) found that 29 of the 48 States had one or more statewide testing programs in the elementary grades, in the secondary grades, or in all grades. In 17 of the 29 States one or more programs were sponsored by the State departments of education. In 12 of the 29 States the only State programs or services available were those sponsored by colleges and universities.

Three years later, in March 1958, a survey conducted by the Educational Testing Service found that there were 31 testing programs in operation in the 17 States and 2 Territories of the 30 States and 2 Territories responding.

With the passage of the National Defense Education Act of 1958 (NDEA), public and nonpublic secondary schools in all States and Territories may participate in statewide or territorywide testing programs. In those States where State laws permit the testing of pupils in nonpublic schools, tests may be conducted as a part of the overall programs. In States which prohibit the use of State funds for nonpublic schools, the U.S. Commissioner of Education is authorized to provide funds for testing by special contract. The eligible nonpublic schools enrolled almost 190,000 secondary school pupils in 1958–59.

Conant Looks at Testing

DISCUSSING his first recommendation for the American high school today, the counseling system, James B. Conant says that, among their qualifications, counselors "should be familiar with the use of tests and measurements of the aptitudes and achievements of pupils." ". . . the counselor," he says, "should be in close touch with the parent as well as the pupil. Through consultation, an attempt should be made each year to work out an elective program for the student which corresponds to the student's interest and ability as determined by tests of scholastic aptitude, the recorded achievement as measured by grades in courses, and by teachers' estimates."

Quoted from James B. Conant's *The American High School Today* (pp. 44–45), The McGraw-Hill Book Co., Inc., with permission of the author.

Elementary as well as secondary schools have programs of testing, including both the usual teacher-made tests and the standardized variety. Usually the standardized tests used in the lower grades include reading readiness and intelligence tests. Tests of achievement may be either teacher-made or standardized, depending on the purpose for testing and the nature of the achievement to be measured. In some elementary schools careful records of individual progress are kept, including test scores as well as other information. If a testing program is to contribute maximum service, it should extend through both elementary and secondary levels, and there should be articulation between these levels.

For Specific Guidance Purposes

Of course, testing programs developed for specific guidance purposes should be planned to meet the needs of the particular school. There is no one "pat" program which can be recommended as best for all school situations. One of the most serious errors into which a school may fall is to adopt a program of testing wholecloth because it has proved successful in another school. There is no substitute for careful and intelligent planning by the local school in designing a testing program to meet its own needs.

Among the pertinent considerations in planning a testing program for guidance are the types of further schooling or occupations into which the graduates typically go, the range of curricular experience available to

pupils, the socioeconomic characteristics of the pupil population, and the qualifications of the staff to interpret test results.

Let us note the last of these considerations—the qualifications of personnel. There is abundant experience to indicate that the qualifications of available personnel place limits on the kinds of testing which may properly be undertaken. For example, the use of individual intelligence scales has no place in a program of testing unless properly qualified personnel are available to administer and interpret them.

One way to gain an impression of the preparation of persons available for testing programs is to examine the requirements of the States for the certification of counselors. However, such an examination gives an incomplete picture since staff members other than counselors may assist in the program. In the elementary grades of most school systems the testing program is generally carried out by the teachers. The same thing is often true in small schools at the secondary levels.

As of 1957, fewer than half of the States *required* the certified counselor to have specific training in testing, as is indicated by the following summary of certification requirements:

14 States *required* college credit in the general area of understanding (or analysis) of the individual, which usually includes training in testing

6 States *required* credit in tests and measurements

1 State *required* credit in individual intelligence testing

2 States *accepted* (but did not require) credit in the general area of analysis of the individual.

Certification requirements also give a minimum level picture of the training for counselors, since counselor preparation programs in a number of colleges and universities are known to hold to standards well above those reflected in certification requirements. Where then shall schools find men and women professionally competent to man their programs? There are several answers. In the long run, the major source of supply will be the colleges and universities, for there is no apparent shortage of institutions offering training programs in guidance and counseling.

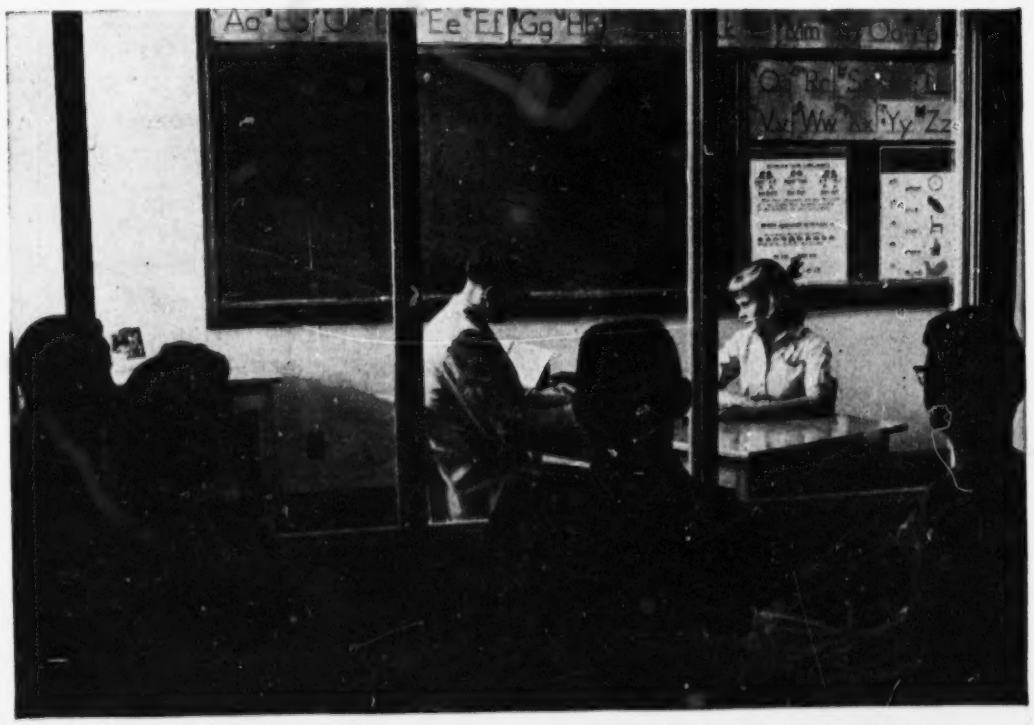
In the short range there are several possibilities. The first source of help is inservice training. Heavy reliance must apparently be placed on efforts by school systems to encourage and to provide it since most of the responsibility of testing programs will be carried by persons already in the schools. A second source of help is the guidance institutes. Under the NDEA, funds are allocated for institutes at selected colleges and universities.

During the past summer 50 guidance institutes were held throughout the country with approximately 2,000 persons in attendance. The full cost of instruction was borne by the Federal Government. The teachers or counselors selected to receive this specialized training in testing and other aspects of guidance were paid a maintenance allowance over and above their tuition.

The very fact that NDEA recognizes testing as a part of the guidance program will stimulate such training. Most States have included in their State plans, submitted in order to be able to receive funds from this part of the NDEA, some requirements for counselors in schools participating in the program. Many of the State plans provide that guidance personnel must be qualified under State certification requirements. While this provision is meant to stimulate more schools to require their staff members to be qualified, it does nothing to raise certification standards. The establishment of standards in State plans will, however, have a salutary effect on States not now having certification requirements.

To Serve All Pupils

While testing programs need to be tailormade for specific purposes, as we noted earlier, this requirement



Summer Guidance Institute, School of Education, University of South Carolina

Throughout the institute, emphasis was placed on the necessity of strict adherence to the instructions and conditions under which a standardized test is administered. In this picture, high school teachers enrolled in the institute are observing the administration of a test, but are unseen by the young girl taking the test. The "one-way" glass in one wall of the testing room and a concealed microphone enable them to see and hear the testing.

need not be interpreted to mean that a program is to be limited to one special purpose. Currently, there is a great resurgence of interest in the identification of the able, an idea that goes back at least as far as Jefferson. The role of tests in the selective process is somewhat newer. Well-considered testing programs, including competent interpretation of test results, can contribute to identification of the able.

But a school testing program established only for this purpose is courting difficulty, especially if "the able" be narrowly and rigidly conceived. A testing program must serve the needs of all pupils in the school which maintains it. This, of course, does not mean that the same test prescription should be administered to all pupils, although there may well be a common core of tests which are administered to all in the interests of administrative efficiency.

Beyond this common core, testing will vary according to differing purposes and special needs of individual pupils. The more able certainly merit consideration—probably more than they have had—but the less able also merit consideration. Two questions will always persist: More able for what? and how can we identify the more able unless we also identify the less able?

With Public Understanding

Not the least of the many problems associated with testing programs is that of interpreting the results to the various school publics. A curious folklore has grown up about testing, and even fairminded and intelligent laymen may well be confused by the claims and counterclaims. In their more relaxed moments many teachers and parents might well wish that the term "IQ" had never been invented. But even if it had not, some other term would probably be equally misunderstood. We perceive what we are prepared to perceive and what we need and want to perceive.

Headlines may build an attack on testing by dramatic coloring of a story of the boy whose test scores were interpreted as indicating poor prospects in college but who nevertheless graduated. Seldom are such questions as the following either considered or answered: What test scores? Who interpreted them? What was the *total* pattern of test evidence? What related information was considered along with test scores? Prospects for success in *what* college? Such questions supposedly do not have reader appeal. Lacking such stimulus to more thoughtful consideration of the special case, the public feels it must defend one grossly misjudged boy by condemning the whole testing program.

There need not and ought not be any secrets surrounding a testing program. Who, after all, is more entitled to complete and accurate information than the parents? The problem is not only one of fact giving concerning the test results but of relating these results to the total developmental pattern of the pupil. Such understanding is the concern of parents, school, and pupil. Testing programs for guidance will stand or fall as they succeed or fail in contributing to their understanding.

Title V—Guidance, Counseling, and Testing; Identification and Encouragement of Able Students

Part A—State Programs:

The act authorizes the appropriation of $15 million for 1958–59 and for each of the 3 succeeding fiscal years for grants to State educational agencies to help them establish and maintain programs of guidance, counseling, and testing.

A State may participate by submitting to the Commissioner of Education through its educational agency a plan which sets forth—

1. A program for testing students in its public secondary schools and, where authorized by law, in nonpublic secondary schools, to identify students with outstanding aptitudes and ability, and the means of testing to be used.

If the State agency does not have authority to pay for testing children in nonpublic secondary schools, the Commissioner of Education shall arrange for testing and pay full cost for 1958–59, and one-half the cost for the 3 succeeding years.

2. A program of guidance and counseling in the public secondary schools to (*a*) advise students of courses best suited to their ability, aptitudes, and skills, and (*b*) encourage students with outstanding aptitudes and ability to complete secondary school, to take courses for admission to institutions of higher education, and to enter such institutions.

Part B—Counseling and Guidance Training Institutes

The act authorizes $6.25 million for 1958–59 and $7.25 million for each of 3 succeeding years and enables the Commissioner to contract with colleges and universities to conduct short-term or regular session institutes for the training of guidance and counseling personnel in secondary schools, and for the training of secondary school teachers preparing to engage in such counseling and guidance. Enrollees from public secondary schools are eligible upon application therefore to receive $75 a week plus $15 a week for each dependent for the period of their attendance at an institute.

By Frank E. Wellman, *Specialist, Program Organization*

ADMINISTRATION *of the* TESTING PROGRAM

Tests can be extremely useful when properly used. They can be harmful when misused.

THE school staff holds the key to the ultimate benefits which may be derived from the testing program. Tests can never be more than a tool to collect the information needed to carry out other functions in the school program. They are merely a means to this end and can never be considered as an end in themselves. The extent to which the school testing program fulfills its purpose is therefore primarily dependent on the administration of the program and the utilization of the test results.

Indiscriminate selection or careless administration of tests and inadequate interpretation of test results or failure to use the results can mean that the testing program is only a waste of teachers' and pupils' time and the taxpayers' money. Where such conditions exist, tests will inevitably be in disfavor with teachers, pupils, and parents. Such a situation can be avoided by the appropriate organization and operation of the testing program as an integral part of the total school program.

Basic educational objectives are related to the development of children, and all professional school staff members should contribute directly to the fulfillment of these objectives. Tests can provide information about pupils which, when properly interpreted, can assist staff members in fulfilling their responsibilities. All staff members should therefore consider the place of test results in the conduct of staff functions and responsibilities to attain the school's objectives and should participate in the planning, organization, and utilization of the testing program if it is to result in maximum benefits.

Most school systems employ an administrative and supervisory staff, an instructional staff, and a special services staff.

The *administrative and supervisory staff*, normally including the superintendent, principals, and instructional supervisors and their assistants, is responsible for determining basic objectives and policies and for providing the facilities needed to make available an environment conducive to good learning and optimum development of each child.

The *instructional staff* has teaching as its primary responsibility for the accomplishment of the objectives of the school.

The *special services staff*, normally including such persons as the school counselor, school psychologist, attendance worker, school nurse, and school social worker, is primarily responsible for providing assistance to individual pupils and to the administrative and instructional staff. The special services staff collects, analyzes, and interprets information about pupils, assists individual pupils in their educational and vocational planning and personal and social adjustment, and fosters pupil physical development and learning in school. To carry out these functions it necessarily works not only with individual pupils but with teachers, parents, and school administrators.

Although each staff member has certain specific functions, all staff functions are interrelated in that they are derived from and contribute to the basic school objectives. Good administrative procedure demands utmost staff cooperation and close coordination of all school functions. The administration, organization, and utilization of the testing program is no exception. In the selection of appropriate tests, efficient administration of the tests, and wise utilization of test results, staff cooperation and coordination are necessary.

Selection of Tests

Since every test in a well-organized testing program should be chosen for a specific purpose, selecting the appropriate tests from the hundreds commercially available is a problem. This problem can be lessened by following the principles described below:

- Define the objectives of the school. For example, each child should read at a level commensurate with his ability. Normal physical growth should be encouraged. Each child should be encouraged to further develop his special skills.
- Determine the areas in which progress toward objectives can be measured with tests. For example, can the reading level be determined? Can physical growth be measured? Can abilities or aptitudes be determined?
- Decide at which grade levels it would be most profitable to use an objective measure. For example, the primary teacher may want immediately certain information which can be obtained from a reading readiness test, mental ability test, and observation. A school counselor, in working with the ninth grade, may need information about pupil abilities, aptitudes, and interests for use in educational and vocational planning. The administrator and instructional supervisors may need information on the general educational development of all pupils at specified grades.
- Select the most suitable tests which are valid, reliable, and practical. The determination of the general validity and reliability of a particular

test is the responsibility of the publisher. (See page 7.) The *practicality* of a test refers to those considerations which determine whether it may fit into a particular school program, such as time limits, ease of marking and scoring, and availability of suitable interpretive aids.

The type of comprehensive analysis necessary in selecting tests is seldom made by one person. In many schools a committee evaluates the tests available and plans the total testing program. Since many staff members may be interested in test results, many may want to take part in the evaluation procedure.

Planning the Test Administration

Once tests have been selected, school administrators should consider a number of details if the program is to be useful. Some testing programs lose much of their effectiveness because of a lack of preplanning. Certainly the following considerations are important:

1. *Orienting the school staff, pupils, and parents.*—The ultimate effectiveness of the testing program is closely related to staff and pupil understanding of the purposes of the tests and the uses of the results. New staff members and pupils need special explanations. Some schools have found it helpful to explain the kinds of tests to be given and their intended purposes to parents and pupils at small group meetings and through printed materials. Such procedures will generally improve the motivation of pupils to put forth their best efforts in taking the tests.

2. *Scheduling the tests.*—Tests should be scheduled in advance so that the disruption of regular classes will be minimized. Test batteries requiring 7 or 8 hours should be given in more than one session. The purpose of a test will determine the appropriate time during the school year. It is desirable to administer some tests in the fall so that teachers and counselors may use the results in planning the pupil's program and in improving instruction. Other tests may be administered in the spring so that the results may be used in evaluating the pupil's progress during the year or in assisting him in selecting a college.

> ***IF MEASUREMENT is to continue to play an increasingly important role in education, measurement workers must be much more than technicians. Unless their efforts are directed by a sound educational philosophy, unless they accept and welcome a greater share of responsibility for the selection and clarification of educational objectives, unless they show much more concern with what they measure as well as with how they measure it, much of their work will prove futile or ineffective.***
>
> ***—E. F. Lindquist, in Preliminary Considerations in Objective Test Construction,* Educational Measurement. *Washington, D.C.: American Council on Education, 1951.***

3. *Administering the tests to the pupils.*—Tests are ordinarily administered by the teachers under the leadership of the school guidance staff. Instructions and time limits must be followed exactly in order that all of the tests may be administered under the same conditions so that local or national norms will be meaningful. Comfortable seating, adequate lighting and ventilation, and efficient provisions for distribution of materials are important in obtaining the pupils' cooperation and best efforts on the test.

4. *Scoring the tests.*—Most standardized tests for the fourth grade and up may be marked on separate answer sheets which may be machine scored or hand scored with special keys. Machine scoring is more accurate than hand scoring and releases the teacher for other more appropriate activities.

5. *Norming the test scores.*—The basic scoring of tests produces only the raw score, that is, the total number of correct responses which the pupil made on the test. Special tables, or norms, indicate a conversion of the raw score to more meaningful scores which provide a comparison of any pupil's test performance with that of the standardization group.

6. *Recording the test scores.*—Test results must be made available in a form understandable to the school staff, pupils, and parents. The name and the form of the test and the test scores are usually recorded in each pupil's cumulative record. Special pupil profile charts or graphs may be constructed to aid in interpretation of scores to the pupil and parent, and class record sheets may be compiled to make test results available to teachers.

A major administrative provision, and one which is frequently neglected, is that of making a pupil's test records from previous school years available when he transfers from one school to another. This practice is beneficial both to the pupil and to his new teachers. At least one State requires that a pupil's cumulative records follow him to his new school.

Utilization of Test Results

Only a few of the possible uses of tests and techniques for analyzing test scores are suggested here, for the nature and extent of test utilization depend on the organization and competencies of the school staff.

Tests results can guide the administrative and supervisory staff in planning an educational program which will be appropriate to pupil needs and will contribute effectively to the achievement of the school's basic objectives. Teachers can provide more effective instruction when they know the strengths and weaknesses of each pupil. For example, Tom was failing in arithmetic. A scholastic aptitude test indicated that he had a high ability and should be doing well in arithmetic. Further testing indicated that Tom had failed to learn to add fractions. With extra help from his teacher Tom soon improved to a more desirable achievement level.

School counselors and other special service staff members constantly need information which they may use in identifying pupil characteristics and situations associated with potential problems and maladjustments. The prevention of pupil problems and maladjustments is a constructive measure and as important as coping with problems after they become evident.

The staff should analyze and interpret test data in conjunction with all available pertinent facts about the pupils. A staff member needs to exercise professional judgment which is based on extensive knowledge of and experience with tests, child development, statistical procedures, educational problems, and guidance techniques. Usually the professional leadership for effective staff utilization of test results rests with the school counselor by virtue of his professional preparation, experience, and the fact that part of his staff assignment is to assist teachers and supervisors in understanding and utilizing test data.

Many different techniques may be employed by staff members in making maximum use of test results. One commonly used technique is the scattergram which indicates the rank of each pupil in a class in two different characteristics. The scattergram, shown in figure 1, based on national standards, shows a comparison of composite school achievement with the scholastic aptitude of each pupil in a group of sixth-grade pupils.

Teachers of these pupils can get several leads from this chart. For example, from scores on the scholastic aptitude test they might conclude that the group has somewhat greater potential to learn than the average population of sixth-grade pupils, since 29 out of the 40 pupils in this group ranked in the upper 60 percent; that is, middle 20 percent or above on a scholastic aptitude test, whereas an average of only 24 might be expected to rank at this level. Furthermore, most of these pupils are achieving at the approximate level of their ability estimated on the basis of their scholastic aptitude; that is, the rank of 31 of the 40 pupils in school achievement was within approximately 20 percentile points of their rank on the scholastic aptitude test. Since school achievement and scholastic aptitude cannot be measured precisely, this agreement between the pupils' ability and their achievement is as close as can be expected. The teacher would give extra attention to Alan, Yvonne, Pearl, Barbara, Ira, Keith, and Thelma, who failed to achieve in school at a level commensurate with their indicated ability.

The causes for underachieving may be many and will vary from pupil to pupil. The scattergram has merely helped the teacher identify pupils not performing at a level reasonably close to their ability—it does not tell her why. More extensive study of the underachievers by the teacher and special service staff members will be necessary to discover probable causes and to determine the needs of these pupils. This is only one example of how the school staff may use test results to identify pupil needs—there are many others.

The school counselor and teachers should interpret the test results to pupils and parents so they may make educational and vocational plans appropriate to the pupil's abilities, aptitudes, skills, interests, and opportunities. They should of necessity make such interpretations primarily in individual counseling conferences with each pupil and his parents. They should present facts clearly and honestly so that pupils and their parents may make decisions intelligently. They should also make every effort to present an accurate and clear picture of the pupil's potentialities. Counselors and teachers cannot assist in answering questions solely on the basis of test results but must consider related information; nor can

Figure 1.—Sixth-grade pupils ranked by fifths on the bases of composite school achievement and scholastic aptitude test scores.

Scholastic aptitude \ School achievement	Lowest 20%	Next 20%	Middle 20%	Next 20%	Highest 20%
Highest 20%	Alan	Yvonne	Pearl	Sandra Edward	Carol Karen Bruce Paul
Next 20%	Keith	Barbara Ira	Dora Henry	Flora Joan David	Ruth Cecil Roger
Middle 20%	Thelma	Hazel	Agnes Mabel Floyd Glenn	Irene Wanda	Vivian
Next 20%	Lyle Neal	Edna Nancy John	Leona	Anita	
Lowest 20%	Grace Opal Marvin	Owen			

Figure 2.—An expectancy table showing the chances of achieving average or better grades in one State university when given the rank attained on a scholastic aptitude test *and* rank in high school graduating class.

An example: a pupil in the middle 20 percent of a scholastic aptitude test and the middle 20 percent in rank in high school graduating class has 1 chance in 3 of achieving average or better grades at one State university.

Scholastic aptitude	Rank in high school graduating class: Lowest 20%	Next 20%	Middle 20%	Next 20%	Highest 20%
Highest 20%	1 in 27	2 in 5	3 in 5	3 in 4	9 in 10
Next 20%	1 in 18	1 in 3	1 in 3	3 in 5	4 in 5
Middle 20%	1 in 7	1 in 5	1 in 3	1 in 2	3 in 4
Next 20%	1 in 46	1 in 6	1 in 7	1 in 2	2 in 3
Lowest 20%	1 in 38	1 in 24	1 in 4	1 in 4	5 in 6

NOTE: This chart was adapted from data assembled at one university and represents only approximations of chances for success in this institution and *should not be interpreted as being applicable to other institutions or other populations of students.*

they help the pupil and his parents make the best decisions without considering the appropriate facts which can be obtained most easily and accurately through the use of tests.

School counselors and psychologists have conducted much research and have developed many techniques to assist parents and pupils in understanding basic data bearing on a problem. For example, when a pupil and his parents must decide whether the pupil should plan to go to college after he graduates, they will want to consider, among other things, whether he has a good chance of succeeding in a particular college. To help them answer this question, the school counselor might again employ the scattergram technique converted to an expectancy table. Figure 2 shows the approximate chance (or expectancy) of a pupil achieving average or higher grades in one State university when his rank on a scholastic aptitude test and his rank in his high school graduating class are known.

By utilizing such techniques as this expectancy table, the school counselor can answer such a basic question as "What are my chances of succeeding in X college?" by saying, "Since you rank in the upper 20 percent of the senior class and in the upper 20 percent on our scholastic aptitude test, the chances are about 9 out of 10 that you will make average or higher grades during your freshman year if you enter X university." On the other hand, if the pupil ranks in the lowest 20 percent of the senior class and in the lowest 20 percent on the scholastic aptitude test, the counselor might say, "Our experience has shown that high school graduates who attained test scores similar to yours and who ranked in the lowest fifth of the graduating class have been successful at X university in only 1 time in 38. This does not mean that you are certain to fail at X university, but it does mean that you will be an exception if you succeed at X university."

Such an interpretation makes test scores and school grades meaningful in solving a specific problem. With these facts the pupil and his parents can plan with some knowledge of the probable outcome of their decision. In any case the responsibility for making the decision rests with the pupil and his parents. The school counselor does not make the decision, rather he collects, organizes, and interprets pertinent information so that pupil and parents can make decisions with more confidence and understanding.

Other school staff members also use test information—school psychologists and school social workers in diagnostic work, instructional supervisors in educational selection and placement, and administrators in curriculum evaluation. The well-organized school testing program is evaluated continuously through systematic research to determine whether it is fulfilling its purposes and to develop more effective means for utilizing the results of the tests.

It is obvious that the effective school testing program demands staff responsibilities which can be carried out only by an adequate staff with the professional competencies needed to make appropriate and prudent utiliza-

tion of test results. Tests can be extremely useful educational tools when properly used—they can be unfair and harmful when misused. Any attempt to conduct a school testing program without staff members specifically prepared to assume the professional responsibilities required is ill advised and will show parsimonious returns for the investment.

The well-organized school testing program in a well-staffed school can contribute significantly to providing maximum educational opportunity for every pupil; to the development of maximum individuality among pupils; to the reduction of educational regimentation and academic conformity; and above all, to the perpetuation of the democratic principle and privilege of every individual to make those decisions which will determine the educational and vocational course of his life.

Definitions of Commonly Used Measurement Terms

ACADEMIC APTITUDE: The combination of native and acquired abilities that is needed for school work; likelihood of success in mastering academic work, as estimated from measures of the necessary abilities. (Also called *scholastic aptitude.*)

ACHIEVEMENT TEST: A test that measures the extent to which a person has "achieved" something—acquired certain information or mastered certain skills, usually as a result of specific instruction.

APTITUDE: A combination of abilities and other characteristics, whether native or acquired, known or believed to be indicative of an individual's ability to learn in some particular area. Thus, "musical aptitude" would refer broadly to that combination of physical and mental characteristics, motivational factors, and conceivably other characteristics, which is conducive to acquiring proficiency in the musical field. Some exclude motivational factors, including interests, from the concept of "aptitude," but the more comprehensive use seems preferable. The layman may think of "aptitude" as referring only to some inborn capacity; the term is no longer so restricted in its psychological or measurement usage.

BATTERY: A group of several tests standardized on the same population, so that results on the several tests are comparable. Sometimes loosely applied to any group of tests administered together, even though not standardized on the same subjects.

DIAGNOSTIC TEST: A test used to "diagnose," that is, to locate specific areas of weakness or strength, and to determine the nature of weaknesses or deficiencies; it yields measures of the components or subparts of some larger body of information or skill. Diagnostic achievement tests are most commonly prepared for the skill subjects—reading, arithmetic, spelling.

EQUIVALENT FORM: Any of two or more forms of a test that are closely parallel with respect to the nature of the content and the difficulty of the items included, and that will yield very similar average scores and measures of variability for a given group.

FACTOR: In mental measurement, a hypothetical trait, ability, or component of ability, that underlies and influences performance on two or more tests. . . . The term "factor" . . . is commonly used to denote the psychological interpretation given to the variable—i.e., the mental trait assumed to be represented by the variable, as verbal ability, numerical ability, etc.

INTELLIGENCE QUOTIENT (IQ): Originally, the ratio of a person's mental age to his chronological age ($\frac{MA}{CA}$) or, more precisely, especially for older persons, the ratio of mental age to the mental age normal for chronological age (in both cases multiplied by 100 to eliminate the decimal). More generally, IQ is a measure of brightness that takes into account both score on an intelligence test and age. . . .

INVENTORY TEST: As applied to achievement tests, a test that attempts to cover rather thoroughly some relatively small unit of specific instruction or training. The purpose of an inventory test, as the name suggests, is more in the nature of a "stock-taking" of an individual's knowledge or skill than an effort to measure in the usual sense. The term sometimes denotes a type of test used to measure achievement status prior to instruction.

Many personality and interest questionnaires are designated "inventories," since they appraise an individual's status in several personal characteristics, or his level of interest in a variety of types of activities.

MATCHING ITEM: A test item calling for the correct association of each entry in one list with an entry in a second list.

MENTAL AGE (MA): The age for which a given score on an intelligence test is average or normal. If a score of 55 on an intelligence test corresponds to a mental age of 6 years, 10 months, then 55 is presumably the average score that would be made by an unselected group of children 6 years, 10 months of age.

MULTIPLE-CHOICE ITEM: A test item in which the examinee's task is to choose the correct or best answer from several given answers, or options.

NORMS: Statistics that describe the test performance of specified groups, such as pupils of various ages or grades in the standardization group for a test. Norms are often assumed to be representative of some larger population, as of pupils in the country as a whole. Norms are descriptive of average, typical, or mediocre performance; they are not to be regarded as standards, or as desirable levels of attainment. Grade, age, and percentile are the most common types of norms.

OBJECTIVE TEST: A test in the scoring of which there is no possibility of difference of opinion among scorers as to whether responses are to be scored right or wrong. It is contrasted with a "subjective" test—e.g., the usual essay examination to which different scorers may assign different scores, ratings, or grades.

Reprinted with permission from A Glossary of 100 Measurement Terms, *by Roger T. Lennon, Director, Division of Test Research and Service, World Book Co.*

PERCENTILE RANK: The percent of scores in a distribution equal to or lower than the score corresponding to the given rank.

PERSONALITY TEST: A test intended to measure one or more of the nonintellective aspects of an individual's mental or psychological makeup. Personality tests include the so-called ***personality inventories*** or ***adjustment inventories,*** . . . which seek to measure a person's status on such traits as dominance, sociability, introversion, etc., by means of self-descriptive responses to a series of questions; ***rating scales*** . . . which call for rating, by one's self or another, of the extent to which a subject possesses certain characteristics; ***situation tests*** in which the individual's behavior in simulated life-like situations is observed by one or more judges, and evaluated with reference to various personality traits; and ***opinion or attitude inventories.*** . . . Some writers also classify interest inventories as personality tests.

POWER TEST: A test intended to measure level of performance rather than speed of response; hence one in which there is either no time limit or a very generous one.

PROFILE: A graphic representation of the results on several tests, for either an individual or a group, when the results have been expressed in some uniform or comparable terms. This method of presentation permits easy identification of areas of strength or weakness.

RELIABILITY: The extent to which a test is consistent in measuring whatever it does measure; dependability, stability, relative freedom from errors of measurement. . . .

REPRESENTATIVE SAMPLE: A sample that corresponds to or matches the population of which it is a sample with respect to characteristics important for the purposes under investigation—e.g., in an achievement test norm sample, proportion of pupils from each State, from various regions, from segregated and non-segregated schools, etc.

STANDARDIZED TEST (STANDARD TEST): A systematic sample of performance obtained under prescribed conditions, scored according to definite rules, and capable of evaluation by reference to normative information. Some writers restrict the term to tests having the above properties, whose items have been experimentally evaluated, and/or for which evidences of validity and reliability are provided.

SURVEY TEST: A test that measures general achievement in a given subject or area, usually with the connotation that the test is intended to measure group status, rather than to yield precise measures of individuals.

TRUE-FALSE ITEM: A test question or exercise in which the examinee's task is to indicate whether a given statement is true or false.

VALIDITY: The extent to which a test does the job for which it is used. Validity, thus defined, has different connotations for various kinds of tests and, accordingly, different kinds of validity evidence are appropriate for them. For example:

(1) The validity of an achievement test is the extent to which the content of the test represents a balanced and adequate sampling of the outcomes (knowledge, skills, etc.) of the course or instructional program it is intended to cover (***content, face, or curricular validity***). It is best evidenced by a comparison of the test content with courses of study, instructional materials and statements of instructional goals, and by critical analysis of the processes required in responding to the items.

(2) The validity of an aptitude, prognostic, or readiness test is the extent to which it accurately indicates future learning success in the area for which it is used as a predictor (***predictive validity***). It is evidenced by correlations between test scores and measures of later success.

(3) The validity of a personality test is the extent to which the test yields an accurate description of an individual's personality traits or personality organization (***status validity***). It may be evidenced by agreement between test results and other types of evaluation, such as ratings or clinical classification, but only to the extent that such criteria are themselves valid.

The traditional definition of validity as "the extent to which a test measures what it is supposed to measure," seems less satisfactory than the above, since it fails to emphasize that the validity of a test is always specific to the purposes for which the test is used, and that different kinds of evidence are appropriate for appraising the validity of various types of tests.

Validity of a test *item* refers to the discriminating power of the item—its ability to distinguish between persons having much and those having little of some characteristic.